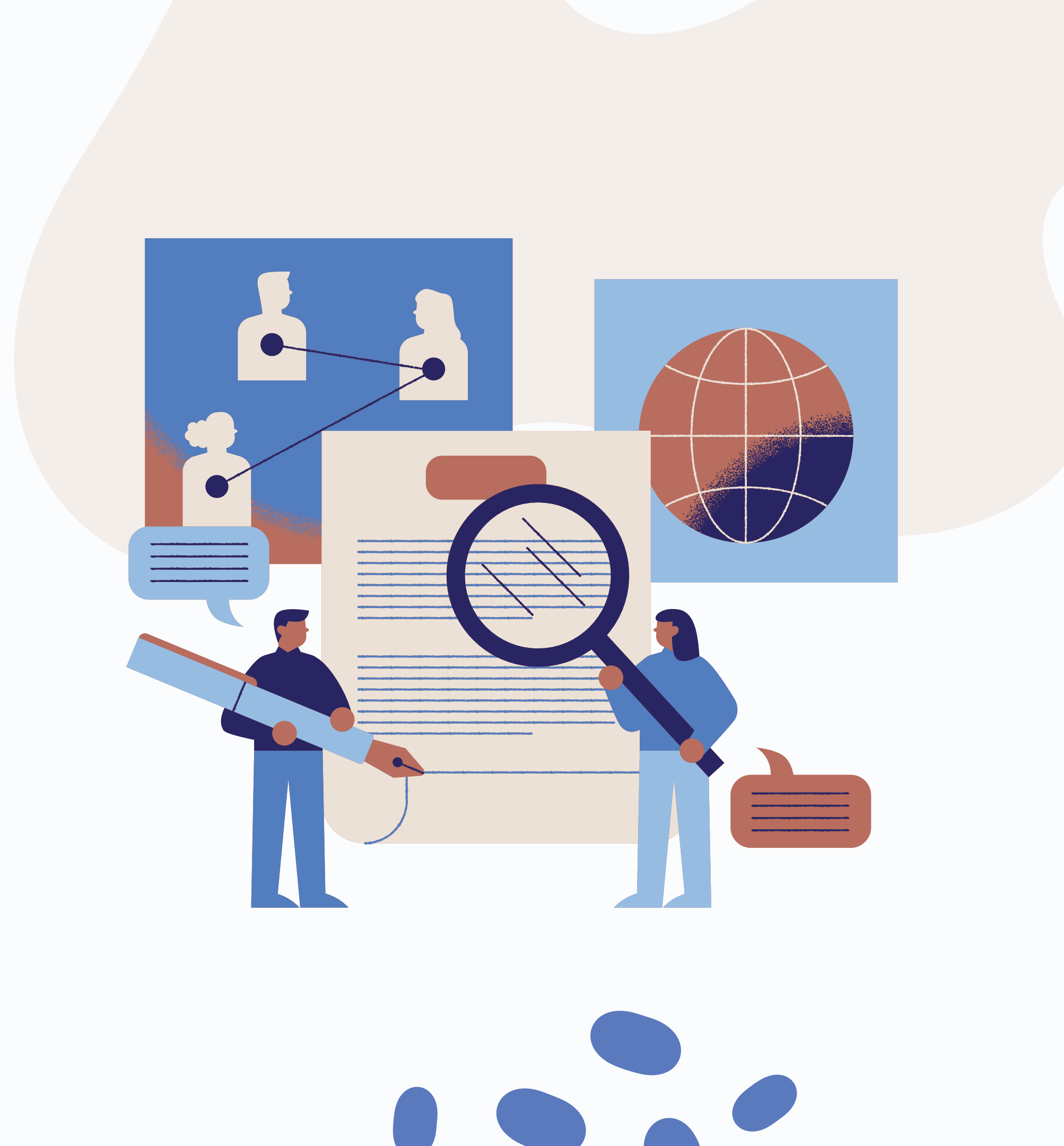


# Data Analysis

Technical Test 1

Sania Rizka Ramadhani



# Tahapan Analisis



01

## Memahami dataset

Memeriksa jumlah dan variasi data pada tiap-tiap kolom. Mengeksplorasi lebih jauh hubungan antar-tabel.

02

## Analisis Insight Data User, Kartu, dan Transaksi

Menemukan insight dari data user, transaksi, dan kartu, seperti mencari frekuensi transaksi per bulan, total transaksi sukses vs gagal, identifikasi user :yang bertransaksi

03

## Membuat Dashboard Interaktif

Memahami pola perilaku pengguna dari waktu ke waktu

# Dataset

## Data Pengguna

id  
current\_age  
retirement\_age  
birth\_year  
birth\_month  
gender  
address  
latitude  
longitude  
per\_capita\_income  
yearly\_income  
total\_debt  
credit\_score  
num\_credit\_cards

## Data Kartu

id  
client\_id  
card\_brand  
card\_type  
card\_number  
expires  
cvv  
has\_chip  
num\_cards\_issued  
credit\_limit  
acct\_open\_date  
year\_pin\_last\_changed  
card\_on\_dark\_web

## Data Transaksi

id  
date  
client\_id  
card\_id  
amount  
use\_chip  
merchant\_id  
merchant\_city  
merchant\_state  
zip  
mcc  
errors

Sebelum melakukan analisis , dilakukan tahapan preprocessing yang meliputi pengecekan missing value dan data duplikat pada ketiga tabel. Selain itu juga dilakukan perubahan format tipe data tanggal yang belum sesuai ke dalam format '%Y-%m-%d' yang valid untuk SQL, seperti kolom 'acc\_open\_date' dan 'expires' pada tabel *card* , serta kolom 'date' pada tabel *transaction*. Tahapan ini menggunakan bahasa pemrograman Python.

# Memahami Dataset

Data yang telah diproses selanjutnya dieksplorasi menggunakan SQL. Ketiga tabel dalam MYSQL yang siap digunakan adalah `user_profile`, `transaction`, dan `card`. Beberapa informasi yang diperoleh antara lain:

## Tabel `user_profile`:

- Terdapat **2000 user**, terdiri dari 984 laki-laki dan 1016 wanita
- Berdasarkan rentang kategori skor kredit FICO, sebanyak 81 pengguna termasuk kategori High Risk, 348 pengguna Medium Risk, 931 pengguna Good, 474 pengguna Very Good, dan 166 pengguna kategori Excellent.

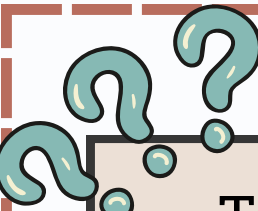
## Tabel `transaction`:

- Data yang diberikan mencatat transaksi pada rentang waktu **1 Januari 2020 - 3 November 2021**.
- Terdapat **2.324.275** transaksi dalam periode tersebut
- Transaksi yang tercatat pada tabel terdiri dari dua jenis, yaitu transaksi melalui **swipe dan online transaction**

## Tabel `card`:

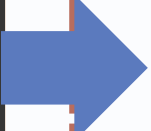
- Pada tabel `card` **tercatat sebanyak 6.146 kartu**, yang dimiliki oleh 2.000 pengguna. Hal ini menunjukkan bahwa data pengguna pada tabel ini konsisten dengan jumlah pengguna pada tabel `user_profile`.
- Sebanyak 5000 kartu diantaranya memiliki chip, sedangkan **646 tidak memiliki chip**.
- Urutan jumlah kartu dari yang terbanyak adalah kartu **debit** (3.511), diikuti kartu **kredit** (2.057), dan kartu **debit (prepaid)** sebanyak 578.
- Brand kartu terdapat 4 jenis, yaitu **AMEX, Mastercard, Visa, dan Discover**

# Insight yang ditemukan



Terdapat 838 user yang tidak melakukan transaksi dari Januari 2010 - November 2011

- Dari 838 pengguna tersebut, sebanyak 43 termasuk kategori high risk dan 176 termasuk kategori medium risk
- Jika dikategorikan berdasarkan usia saat ini terhadap batas usia pensiun, terdapat 792 pengguna pada usia produktif yang tidak melakukan transaksi.



Tabel tersebut menyajikan gambaran spending, refund, jumlah transaksi, serta frekuensi transaksi.

Pada kolom amount, terdapat nilai negatif yang diinterpretasikan sebagai transaksi refund. Oleh karena itu, total transaksi dipisahkan menjadi tiga kategori utama: spending, refund, dan transaksi bersih.




Selanjutnya, data dikategorikan berdasarkan rentang waktu untuk mengidentifikasi pola transaksi dan aktivitas refund yang terjadi pada periode tertentu.

## Banyaknya Transaksi

2.324.725

Frekuensi transaksi terbilang cukup besar, dengan rata-rata transaksi per harinya mencapai 3459.

Sementara itu, untuk mengetahui tingkat keramaian transaksi dalam periode waktu tertentu (dini hari, pagi, siang, sore, malam), distribusi transaksi dapat dilihat sebagai berikut:

Result Grid    Filter Rows: <input type="text"/>   Export:  Wrap Cell Content: 					
jam	total_spending	total_refund	jumlah_transaksi	frekuensi_transaksi	
00:00 - 06:00	7861217.52	-848663.00	7012554.52	134642	
06:00 - 09:00	17271117.22	-1999106.93	15272010.29	440621	
09:00 - 12:00	19763979.70	-1850746.93	17913232.77	466770	
12:00 - 15:00	23787410.11	-2768259.29	21019150.82	480115	
15:00 - 18:00	20834757.82	-2290013.22	18544744.60	388250	
18:00 - 24:00	23841563.39	-2488578.30	21352985.09	413877	



# Analisis Transaksi

Dalam dataset ini terdapat sejumlah transaksi yang mengalami kegagalan. Jika dilihat berdasarkan jenis transaksi, baik fisik maupun online, distribusi error menunjukkan pola sebagai berikut:

	use_chip	total_error
►	Swipe Transaction	30883
	Online Transaction	5796

Pada transaksi fisik (swipe transaction), error terbanyak disebabkan oleh Insufficient Balance dengan 20.935 transaksi, diikuti oleh Bad PIN sebanyak 5.622 transaksi, dan Technical Glitch sebanyak 4.024 transaksi.

Sementara itu, pada transaksi online, error terbanyak juga disebabkan oleh Insufficient Balance dengan 1.972 transaksi, diikuti oleh Bad Card Number sebanyak 1.247 transaksi, Bad CVV sebanyak 1.020 transaksi, serta Bad Expiration sebanyak 978 transaksi.

Rata-rata error per hari adalah 54 transaksi

- Jika ditinjau berdasarkan periode waktu per jam, jumlah error tertinggi terjadi pada pukul 07.00 dengan total 2.901 transaksi.
- Selain itu, rentang waktu pukul 11.00 hingga 16.00 juga menunjukkan jumlah error transaksi yang relatif tinggi.
- Apabila ditinjau berdasarkan rentang waktu dapat dilihat sebagai berikut:

	rentang_jam	jumlah_error
►	12:00 - 15:00	7813
	06:00 - 09:00	6818
	18:00 - 24:00	6702
	15:00 - 18:00	6629
	09:00 - 12:00	6390
	00:00 - 06:00	2327



# FURTHER EXPLORATION

Analisis yang telah dilakukan masih memiliki keterbatasan, dan masih terdapat ruang untuk eksplorasi lebih lanjut. Selanjutnya, hasil analisis ini akan disajikan dalam bentuk dashboard interaktif untuk mempermudah pemantauan dan interpretasi data.

Selain mengelola periode transaksi berdasarkan waktu dan mengeksplorasi error pada transaksi, dapat dilakukan analisis demografi pengguna, seperti pemahaman sebaran usia pengguna serta distribusi lokasi transaksi. hal ini akan memberikan wawasan lebih mendalam mengenai profil pelanggan dan pola perilaku transaksi.



# THANK YOU

Feedback and suggestions are welcome via the contact  
information below

github

[github.com/saniarizka/data-analyst-user-behavior-test](https://github.com/saniarizka/data-analyst-user-behavior-test)

Email

saniarizka21@gmail.com

